Designing, Implementing and Evaluating Recommender Systems on Mobile Services Using Big Data Analytics through Apache Mahout

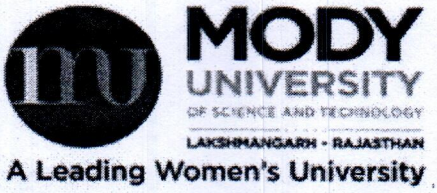
Synopsis of Pre-Submission Report
for the Award
Degree of

DOCTOR OF PHILOSOPHY

IN ENGINEERING

OF

MODY UNIVERSITY OF SCIENCE AND TECHNOLOGY



SUPERVISOR
Prof. V.K Jain
Dean,CET, Mody University
Lakshmangarh, Sikar (Raj.)

Submitted by
Neetu Singh
Enrol. No.: 110753

OCTOBER, 2018
College of Engineering and Technology
MODY UNIVERSITY OF SCIENCE & TECHNOLOGY
(Formerly, Mody Institute of Technology and Science)
Lakshmangarh, Sikar-332311
(Rajasthan), India

MODY UNIVERSITY
LAKSHMANGARH • RAJASTHAN
A Leading Women's University

COLLEGE OF ENGINEERING AND TECHNOLOGY (CET)
Lakshmangarh (Sikar), Rajasthan 332 311, India
Ph. No. : (01573) 225001-12 • Fax : (01573) 225042
email : dean.cet@modyuniversity.ac.in • www.modyuniversity.ac.in

DECLARATION

I, Neetu Singh, hereby declare that the work presented in this thesis entitled "***Developing, Implementing and Evaluating Recommender Systems on mobile Services Using big data analytics through Apache Mahout***" has been carried out by me under the supervision of Prof. Anil Kumar for the award of Degree of Doctor of Philosophy in Engineering, Department of Computer Science Engineering, Mody University of Science & Technology, Lakshmangarh, (Rajasthan) India. To the best of my knowledge, the work is original and has not been submitted elsewhere for any other academic attainment.

It is further declared that I have fulfilled all the requirements laid down by Mody University of Science & Technology, Lakshmangarh for the submission of the report.

Neetu Singh
Enrol. No: 110753

MODY UNIVERSITY
LAKSHMANGARH · RAJASTHAN
A Leading Women's University

COLLEGE OF ENGINEERING AND TECHNOLOGY (CET)
Lakshmangarh (Sikar), Rajasthan 332 311, India
Ph. No. : (01573) 225001-12 • Fax : (01573) 225042
email : dean.cet@modyuniversity.ac.in • www.modyuniversity.ac.in

Date:

CERTIFICATE

This is to certify that **Neetu Singh (Enrol. No. 110753)** has successfully completed the research work, entitled ***"Developing, Implementing and Evaluating Recommender Systems on mobile Services Using big data analytics through Apache Mahout"*** and has submitted the report as a partial fulfillment for the award of the degree of Ph.D. in Engineering, Mody University of Science & Technology, Lakshmangarh. To the best of my knowledge, the work is original and has not been submitted elsewhere for any other academic attainment.

It is further certified that **Neetu Singh** has fulfilled all the requirements laid down by the Ph.D. Ordinance, Mody University of Science and Technology, Lakshmangarh for the submission of the thesis.

SUPERVISOR

11/10/18

Prof. V.K Jain
Dean,CET,
Mody University
Lakshmangarh, Sikar (Raj.)

ACKNOWLEDGEMENT

First of all, I wish to express my sincere gratitude to my supervisor Dr. V.K Jain, Professor & Dean, College of Engineering and Technology, Mody University Lakshmangarh for supervising the work with invaluable inputs. I am particularly indebted to him for hisvaluable suggestion and comment to develop sound logical sense in scientific pursuits which has helped me in developing a broad perspective. Each conversation with him added invaluable aspects to the implementation and broadened my perspective. It was a great pleasure for me to have a chance of working with him. I have exploited each and every minute of his valuable times for preparing my Ph.D. thesis. I am really indented to my supervisor for his noble thoughts.

I wish to express my gratitude to Dr. N. K. Joshi, Professor & Head, Department of Nuclear Science and Technology Mody University of Science and Technology, Lakshmangarh for his kind help and necessary support as and when required.

The successful culmination of this work could not be possible without the support and blessing of my loving parents and parents-in-law. I thank my husband Major Gaurav Singh for his patience, encouragement and cooperation during the entire period of my Ph.D. work. I extend my special thanks to Dr. Sangeeta Jhajharia for supporting me unconditionally during entire Ph. D phase. Without all of them, I would have been greatly incapacitated. Further, I would like to give thanks and regards to my loving son Kritin & my younger brother Aditya for their innocence which refreshed me during my involvement in this work.

My thanks and regards are also due to the once whom I have inadvertently forgotten to mention but have helped me directly or indirectly towards this work successfully.

Last, but not the least, I thank the Almighty God, for successful completion of my Ph.D thesis work.

NEETU SINGH

ABSTRACT

Big data is an emerging research paradigm. Big data age has come now. As the information technology spreads fast, the communication is now more of digital. WhatsApp, Facebook, twitter ,linkedin , google maps and online shopping have become a daily routine affair of a common man. It ripens the need of storing, processing and exploration of big data e.g. the MapReduce paradigm. With time it is recognised that it is necessary to further develop platforms and applications that can harness these technologies in order to achieve meaningful insight and to make more accurate business judgments. One of impressive move in such direction is Recommendation systems; it is something that identifies user interests and preferences based on the knowledge of the user's behaviour and on the item's characteristics that they have used.

Rapid increase in cellular network users has been noticed on the daily basis. With the growing cellular trends and to be a survivor in the cut throat competition amongst cellular network providing companies, it is the demand of the hour to provide a consumer with a convenient mode of communication .A mode that they thoroughly enjoy and give them feel that their chosen network is best in each and every aspect.As a network providing company, the main concern is how to increase business and network while providing best services to the user. For all these services users have to pay. There are the tariff, SMS, data and roaming plans which user must subscribe to meet his needs. Companies generally inform their users about such plans through SMS. But the crude fact is, it's very annoying for the customers to receive all these messages which are of no use for them. Today an individual is so occupied by innumerable tasks and timelines that he feels that involving him in all this is just a waste of time, How about a recommendation system that hits his interest specifically and stops all other irritating messages.Every competitive company wants to attract new customers, while retaining the exciting ones. In short, they want to churn out customers at any cost. This concept become more important in today's environment where customers have a numerous opportunities for fulfilling their needs. All they want is best services in minimum cost. And if they get it from another provider, they change their network without giving a second thought. In such cases, providing plans that suits to customer interest is a great idea.

Our work presents user based and item based recommenders which are capable of enhancing the user experience towards its cellular network by utilizing the Mahout Big data analytics library. It keeps user interest at topmost priority and stops superfluous messages which are not of any use to them.With time CF has become a striking subject for researchers because of its competency to handle information overload efficiently. Many techniques and different approaches have been developed to generate an effective recommendation. This is interesting as different application's scenarios could have a fittest solution. Considering this, finally we have discussed the different techniques that are capable to enhance the efficiency of RS and proposed similarity fusion technique to provide better prediction accuracy.

TABLE OF CONTENT

		Page No.
Acknowledgement		**i**
Abstract		**ii**
List of Figures		**vii**
List of Tables		**viii**
Chapter 1.	**Introduction**	**1-5**
1.1	Introduction	1
1.2	Motivation	2
1.3	Research Gap	4
1.4	Objectives of the Research Work	5
1.5	Overview of the Thesis	5
Chapter 2.	**Recommender Systems**	**6-11**
2.1	Introduction	6
2.2	Recommender system dimensions	6
2.3	Recommender Process	6
2.3.1	Data Creation	7
2.3.2	Recommendation Approaches	8
2.3.3	Predictions and Recommendations	10
3	Summary	10
Chapter 3.	**Evaluating performance in recommender System**	**12-16**
3.1	Introduction	12
3.2	Evaluation Metrics	12
3.3	Error-based metrics	13
3.4	Precision-based metrics	14
3.5	Experimental setup	14
3.6	Summary	15
Chapter 4.	**Design & implementation of RWYW mobile recommender using Tanimoto, Euclidien& Pearson correlation similarities**	**17-27**
4.1	Introduction	17
4.2	Proposed Approach	19
4.2.1	Input	19

	4.2.2 Item Content	20
	4.2.3 User Profile	21
	4.2.4 Batch Processing and Model Repository	21
4.3	Recommendation and output	22
4.4	Algorithm	22
4.5	Results and Discussion	23
4.6	Evaluation	23
4.7	Summary	26
Chapter 5.	**Design & implementation of item recommender for mobile plans using Tanimoto, Euclidien& Pearson correlation similarities**	28-34
5.1	Introduction	28
5.2	Mobile plans item recommender architecture	29
	5.2.1 Phase1	29
	5.2.2 Phase2	29
	5.2.3 Phase3	30
	5.2.4 Phase4	30
	5.2.5 Phase5	31
5.3	Results and Discussions	31
	5.3.1 Evaluation	32
5.4	Summary	34
Chapter 6.	**Efficiency Booster For Recommendation System**	35-
6.1	Introduction	35
6.2	Item recommender Architecture using Similarity fusion	35
6.3	Experimental Results	38
6.4	Performance Analysis	39
6.5	Summary	40
Chapter 5.	**Conclusion and Scope ofFuture Work**	41-42
List of Publications		
Mapping of Research Work Objectives		
References		

LIST OF FIGURES

Fig. No.	Description / Title of Figures/Graphs	Page No.
2.1	Recommendation Process	7
4.1	Yearly increase in mobile devices	18
4.2	Proposed RWYW Architecture	19
4.3	RMS error Verses similarity	24
4.4	precision Verses similarity	25
4.5	nDCG Verses similarity	25
4.6	Fallout Verses similarity	25
5.1	Architecture of mobile plans item recommender	30
5.2	Similarity Vs RMSE	32
5.3	Similarity Vs Precision	33
5.4	Similarity Vs Fall out	33
5.5	Similarity VsnDCG	33
5.6	Comparison Graph of User based & Item based Recommender	34
6.1	Architecture of mobile plans item recommender using similarity fusion	36
6.2	Similarity fusion effect on RMSE	39
6.3	Similarity fusion effect on Precision	39
6.4	Similarity fusion effect on Fall out	40
6.5	Similarity fusion effect on nDCG	40

LIST OF TABLES

Table No.	Description/Title of Tables	Page No.
2.1	User-Item matric	9
4.1	Airtel mobile plans	20
4.2	Neighbourhood size impact on execution time	26

Chapter – 1

Introduction

1.1 Introduction

New items that are of user interest have been suggested on the basis of user data utilisation after applying appropriate algorithms with the help of tools and techniques known as Recommender System. It enhances user experience by helping them to take a decision by providing various items based on preference value while doing online shopping, e-commerce, browsing data on internet and during content search on the web. Preference values are based on user's past history or on the other user's experience. There are dual ways to acquire information ,one is explicitly such as collecting user's ratings or other one is to gather it implicitly [1,2] like by observing user's behaviour, like which songs the user hears, kind of applications user downloads, which books user likes along with his personal information such as sex, nationality, age, name. Sometimes, recommender systems may also use public media information such as friend's count, friend following, tweet counts and posts. With advancement of technology, the data collected via GPS, RFID is of great help [3].It is critically important that recommendation system keeps balance between various factors like innovation, consistency and accuracy. The key to RS is Collaborative Filtering (CF). CF considers human ways of making decisions. With the popularity of RS, it is clear that it handles diversity in areas well. Movie recommenders [4,5] are one of the most current and common topic for many research papers. However, television [9], music [6,7,8], documents [12,13,14], books [10,11], e-Learning [15], e-commerce [16], marketing applications [17] and searching web [18] are other fields that have been explored by RS. IDC's Digital Universe study had conducted a survey stating that the world information doubles itself in every two years. The forecasts have clearly pointed out that it will be reaching 40 ZB by 2020 [19]. Moreover, IBSG (The Cisco Internet Business Solution Group has predicted that 25 billion devices will be internet connected latest by 2015.This increase is further estimated up to 50 billion by 2020[20]. Every day the numbers of the mobile users of cellular networks are increasing.So it's no longer a question whether mobile marketing important, we know it is! It's now a matter of using the statistics to understand how users behave when using different types of mobile devices and what their preferences are [21]. As the numbers of users are in millions, the data generated is really huge. This increase in data and such vast datasets often referred as "Big Data". In growing distributed systems, there is question on traditional

distributed system's scalability, storing, retrieving and analyse of huge data. Hence the information stored in data warehouses for such solutions are no longer statistically reliable. So, there is need of brand new architectural designs which support big data's scalability requirements and match its parallel processing speed. The implementation of data analytics on big datasets is generally named as big data analytics. Recently, machine learning and data mining techniques are incorporated into variety of industries for making more benefits, bear fewer risks and to optimise their respective business. In other words, big data analytics is considered as necessary requirement for such organizations which starves for competitive benefits. IBM's Global CIO study of 2011 had already focused on the big data analytics along with business intelligence for next five years [22]. Recently, mobile marketing has been recognized as a leading field that explores and use big data analytics to meet customer expectations, to be in competitive environment successfully and to bear economic pressure gracefully.

1.2Motivation

The RS are designed to provide recommendations best suited to the users. There is no doubt; they are the boon to commercialization sector.As e-commerce grows, online websites grow and hence grows the set of choices for the users. It is hard for the user to choose from large set of choices. In this situation, recommender lays an important role by providing set of items of user interest. So, RS was considered as solution to this problem that facilities user with set of appropriate choices. The following can be considered as functional objectives for RS

1. **Gain more sales Profit**-Main target of RS is to increase profit by the means of increase in sales. It attracts user to buy more products by suggesting the products of their choice. Sometimes people end up buying three items instead of two.

2. **Diversity Coverage**-RS may perhaps be able to provide those recommendations, which otherwise are not easy to find by user like unpopular movie through movie recommender. It is understood that service provider most likely doesn't want to advertise unpopular movie without RS.

3. **Enhance experienceof User**-If RS provides good recommendations and easy access to user, it in turn increases the user satisfaction. It then becomes a interface, a user cherish during e-commerce, searching and browsing. Also, it rises the rate of acceptance of recommendations.

4. **Hold Users-** A website should hold the user somehow due to which users must have faith on their visited website. This will be achieved by treating a customer as a a valuable customer by recognising the old customer, using his previous interaction history and improving the user model with each and every visit. Finally, RS is capable to provide best recommendation that match the user preference.

5. **Deliver desired Information-** Service provider may use the gathered information to re – use it to achieve the improvised RS, like in item stock management or creating new data sets according to user demands. For e.g. A traveller RS may add new district or city if one or more user inquires about it.

Other than objectives of RS, there are certain tasks that are associated with it. Some of them are opportunistic [23]. The following are some opportunistic tasks of RS:-

- **Predict appropriate items-** It should advise the list of some items with preference value to use. Preference value tells that what are the chances of the user to like the suggested item, may be in form of some ratings or score between 1 to 5.
- **Predict all appropriate items-**It should suggest all the possible items that can satisfy the choice of user. This is not at all possible for RS that works on fat data sets. For this task, RS requires more rules and explanation. This task is most beneficial for mission critical RS.
- **Generate versatile sequence** - Instead of generating a single suggestion of interest, RS should provide the versatile result, for e.g. if it recommends a chain of TV channels, then it may also recommend music tracks and interesting novel according to the mood of user.
- **Suggest bunch** - IT is better to provide a complete bunch of item that fit together very well. Say, for traveller recommender, RS can bunch destinations, main places to visit, accommodation and travelling information, so that the user can select multiple choices in one travel plan.
- **Valuable browsing-** Sometimes, the user may just browse through network for information only, and not for purchasing any item or product. In such cases RS task is to suggest item of user interest that can fit into user's perspective. So that, his

browsing will be converted intovaluable browsing and may be the user end up by adding something to cart in case of online shopping.

- **Stand on user's testing criteria-** In some cases, user wants to test the recommender that how good is it. So, user keeps on testing the recommender. Some recommenders have the explicit feature to let the customer to perform testing on it.
- **Up gradation-** This task basically depends on user, if user provides or updates his personal preference like his likes/dislikes, the recommender can improvise itself with this information. In turn this information improves user model of RS and is capable for providing much better recommendation.
- **Collect other user's feedback-** Feedback is very crucial for RS to work intelligently and smoothly. Like, in the case of car RS, the feedback of user's who have already purchased car is very important since on the basis of this only, RS calculates the rating for that particular car.
- **Intelligent Analysis of Information-** In web based RS, some user may take advantage by providing wrong feedback or ratings to influence other users for purchasing some different product to promote it. So, one of the main tasks of RS is to filter and intelligently use information.

1.3Research Gap

1. **Diversity**
    - RS must be more focussed towards diversity of recommendations because higher the diversity, better the chances for user to find his desired item. In many cases, the user is using the recommender as a facts finding tool. This research topic is still in an early stage, and there is a need to characterize the nature of this "diversity".
2. **Conversational RS**
    - Conversational RSs is an open research area as it can be further improved by implementing learning capabilities that can optimize not only the recommend items but also how the dialogue between the user and the system must unfold in all possible situations.
3. **Mobility**
    - Recommenders designed to operate under mobility i.e for mobile

Mobile computing is a most natural platform for personal computing. Many recommendation requests are likely to be made when the user is on the move.This ripens the need of "mobilizing" the user interface and to design solutions that can efficiently use the limited resources of the mobile devices.

4. **Cold Start problem**
   - The recommender systems also suffer from issues like cold start, sparsity and over specialization. Cold start problem is that the recommenders cannot draw inferences for users or items for which it does not have sufficient information.

1.4 Objectives

- To design and implement user based mobile recommender based on Tanimoto, Euclidean distance and Pearson correlation similarities.
- To design and implement item based mobile recommender based on similarities measures such as Tanimoto, Euclidean distance and Pearsoncorrelationssimilarities.
- To evaluate the quality of mobile recommender using standard available methods.
- To design and implement efficiency booster technique for mobile recommender

1.5 Overview of the Thesis

The thesis is presented trough five chapters and a brief outline of each chapter is given below: Chapter 1 discusses the background, concepts, motivation and objectives behind the current research work.Chapter 2 discusses about Recommender systems withits dimensions, approaches and process along with various similarity algorithms, they used. Whereas, Chapter 3 focuses on performance evaluation of recommender systems. Chapter 4 covers design, implementation and result analysis of user based "RWYW "recommender system. Chapter 5 covers design, implementation and result analysis of item based recommender system for mobile plans. Chapter 6 covers the similarity fusion concept and a method to boost efficiency of recommender System.

Finally, Chapter 7 presents a summary of the significant results and conclusion of the thesis, together with some suggestions for future work in this area.

Chapter – 2

Recommender Systems

2.1 Introduction

Throughout the internet there are web sites, applications, and systems being built with a focus on user interaction and data. The users of these systems expect to be introduced to new content, to be recommended content that their friends like, and want interfaces through which they can submit feedback to improve these recommendations. Recommender systems (RSs) are the tools and techniques that address these demands by utilizing user data and algorithms to suggest new items that will be of use to users. A RS can provide suggestions for products to buy, books to read, places to eat, or movies to watch.

2.2 Recommender system dimensions

The definition, components, and dimensionality help us describe and understand RSs. This leads to a discussion of the specific implementation of commonly used RS algorithms that are evaluated and implemented in this thesis. Every RS attempts to predict items that a user will find most relevant and useful. While this concept is common across all types of RSs, the manner by which a RS calculates relevance and usefulness varies. The amount and type of available data about RS components such as users, items, and preferences often dictate how this relevance and usefulness is calculated and ultimately impacts a RS algorithm selection. When data about a user and their preferences are lacking, a Non-Personalized RS can be an appropriate algorithm selection. A Non-Personalized RS algorithm will rely on the overall data about popular items amongst all users and generate recommendations such as a Top-N list of most popular items. Non-personalized recommendation algorithms do not provide personalized or diverse recommendations to different users based on past preferences of users. Instead, the RS assumes an item that is liked by most users will also be liked by a generic user. While not heavily researched, non-personalized algorithms provide a simple and effective interface to provide recommendations to users when they lack previous preferences, also known as the cold start problem.

2.3 Recommendation Process

Recommendation process is a continuous process of gathering information, extract useful information and then apply suitable recommendation approach to make best possible recommendation that fits user's choice.

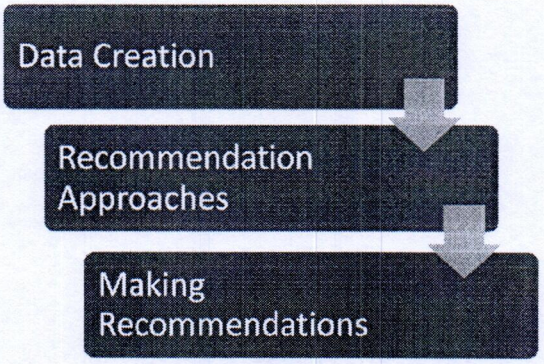


Figure 2.1: Recommendation Process

2.3.1 Data Creation

RSs are information processing systems that dynamically gather a number of data in order to make their recommendations. Since, the data and knowledge sources available for recommender systems is quite diverse, ultimately, either they can be exploited or not depends on the recommendation. Generally, recommendation techniques used in RS are directly proportional to performance of RS. Some techniques use very simple information for the recommendation process like ranking or rating of items. On the other hand, some techniques use advance knowledge like ontological, social and personal constraints with activities associated with them. Data used in RS revolves around users, items and between both of them in the form of transactions. Some values are assigned to an item on the basis of its utility to the user. Positive value for items means that it is of user interest while negative value means that it is of no use for user. It is important to know that when customer is paying for an item, he also bears the cognitive cost of item like searching and monitoring. It is the responsibility to see if the user bears this cognitive cost, even if he is not purchasing item, it must be in the form of useful information and not for irrelevant suggestions. Items are represented using various approaches. One of the simplest ways to represent an item is by assigning it a single ID, whereas in the higher form it is represented by a set of attributes. In RS, the information can be structured in different ways and the selection procedures are based on the recommender approach. Collaborative filtering (CF)is dependent on item ratings provided by user. In contrast Demographic RS is dependent on demographic attributes such as age, sex, profession [24]. RS is a tool that makes use of modelling techniques and creates

user models to produce recommendations [25, 26]. It is a fact that no personalization exits, without a suitable user model. The user model will always play a crucial role. Sometimes, the processing behaviour of user is taken into account, mostly in case of web and traveller recommenders [27].

Transactions: It can be referred as communication between the customer and RS. Transactional data stores important information about RS-customer communication. In some cases, it may also contain the user feedback and rating that he provided to items. Ratings are of two types – explicit and implicit. In explicit, user is asked to give ratings whereas in implicit, system calculates ratings on certain factors. Ratings can be of different forms [28]:-

a. Numeric rating: In the form of some numbers, may be in between 1 to 5, such as movie rating.

b. Opinion rating: It is in the form of agree, disagree, strongly agree and strongly disagree. Mostly, it is collected through questionnaire.

c. Single rating: Just in form of Yes or No, for e.g. 'seen or not seen' in case of movie and 'purchased or not purchased' in case of item.

RS continuously gathers the transactional information and according to its recommendation approach modifies and monitors the output of recommendation process [29].

2.3.2 Recommendation Approaches

For recommendation approach CF algorithm is most common and efficient technology which is used in various fields such as e-commerce and social networking. There are two assumptions for this algorithm:- If two users have same interests, then they both are similar.Say, X and Y are items. If user likes X, and if X and Y are similar to each other than there are bright chances that user also likes Y.The key objective of CF is to calculate the item-user similarity and use it in formation of ranked list which in turn suggests items to user according to his interest.

- **Collaborative filtering process**

CF algorithms are really simple and effective. These algorithms are further sub-divided into two sub classes:-

- User based collaborative filtering – This CF focuses on the user-user similarity, i.e. how much two users are similar.
- Item based collaborative filtering – This CF focuses on the item-item similarity, i.e. how much two items are similar instead of user.

➢ **User-Item matric**

This matric considers item as column and user as row. If there are 'm' users and 'n' items, the 'm x n' matrix is expressed as Rm x n.

Item/user	Item a	Item b	Item c
User 1	5	4	4
User 2	4	3	5
User 3	2	5	3

Table 2.1: User Item Matric

In matrix, the evaluation value for item 'k' from user 'j' is denoted by R (k,j).

➢ **Similarities**

There are various methods that can be used to measure similarity between user and item. Mathematical formulas are there to calculate similarity, but each formula has one thing in common, each works on sum of terms over the set of common user [30]. The following are the various similarities:

- **Cosine-based similarity**:It considers items and their corresponding ratings as vectors and the angle which is formed between them is used for measuring similarity denoted by cos (i, j) and given by:

$$Cos(i,j)=i.j/||i||.||j||.....[a]$$

where "." denotes the dot-product of two items.

Similarity based on Pearson correlation:It calculates the similarity by using ratings of common users for pair of items, diverse from average ratings of these items:

$$sim(i,j) = \frac{\sum_{u \in U}(R_{u,i} - \bar{R}_i)(R_{u,j} - \bar{R}_j)}{\sqrt{\sum_{u \in U}(R_{u,i} - \bar{R}_i)^2}\sqrt{\sum_{u \in U}(R_{u,j} - \bar{R}_j)^2}} \quad \text{....[b]}$$

- **Adjusted similarity based on Cosine**: Sometimes the habits of users affect their ratings. For e.g. some of the users generally give high rating out of their behaviour, while other can give lower rating as a priority. In such cases adjusted cosine based similarity can be used as efficient solutions in which similarity is calculated on the basis of co-rated pair.The similarity between items i and j is expressed by:

$$sim(i,j) = \frac{\sum_{u \in U}(R_{u,i} - \bar{R}_u)(R_{u,j} - \bar{R}_u)}{\sqrt{\sum_{u \in U}(R_{u,i} - \bar{R}_u)^2}\sqrt{\sum_{u \in U}(R_{u,j} - \bar{R}_u)^2}} \quad \text{....[c]}$$

- **Similarity based on Log Likelihood Ratio (LLR):** It is based on statistics which revolve around user or item occurrences. It focuses on the events where item and user overlap and calculate what are the chances that the overlap is genuine or it happened by chance. For e.g. two users are more similar if their preferences match score of six out of fifteen than nine out of seventy five.

2.3.3Predictions and Recommendations:Once we create the recommendation model using any of the similarity discussed above, we are able to predict the rating for any user-item pair.The formula for the predicted rating is given by:-

$$P_{u,i} = \frac{\sum_{\text{all similar items, N}}(s_{i,N} * R_{u,N})}{\sum_{\text{all similar items, N}}(|s_{i,N}|)}$$

- **Content Based Filtering**: This filtering approach uses a set of discrete characteristics of an item to recommend more items with similar properties. It is based on item description and preference provided by user. For example, the music recommendation 'Pandora' uses the properties of a song or the singer in order to produce a station which plays music with similar properties.

2.4. Summary

Along over two decades of research and commercial development, recommender systems have proved to be a successful technology to overcome the information overload that burdens

users in modern online media. The inherent possibility of dealing with diverse sources of information, such as the content of the items and the collaborative and social interactions among users and between users and a system, has enabled the development of rich strategies based on each of these evidences, deriving content-based, collaborative, and social filtering recommendation approaches. Furthermore, as each particular type of recommendation technique has its own limitations and weaknesses, hybrid strategies have been proposed that combine the suggestions generated by different techniques in different ways. The success of ensemble approaches has been recently evidenced in the Netflix prize, where the top classified teams used different forms of recommender ensembles.

There are, however, general limitations remain unsolved, and are still considered as open research problems in the field. Some are particular to specific recommendation techniques while others are independent of any recommendation technique, have been identified in the literature, and deserve special attention by themselves, But, choosing the right recommendation technique is very important for good recommenders

Chapter – 3

Evaluating performance in recommender System

3.1 Introduction

The evaluation of recommender systems has been a major object of study in the field since its earliest days, and is still a topic of on-going research, where open questions remain [31]. Two main evaluation protocols are usually considered [32]: online and offline. In this thesis we focus on offline evaluation.

Drawing from methodological approaches common to the evaluation of classification, machine learning and information retrieval algorithms, offline recommender system evaluation is based on holding out from the system a part of the available knowledge of user likes (test data), leaving the rest (training data) as input to the algorithm, and requiring the system to predict such preferences, so that the goodness of recommendations is assessed in terms of how the system's predictions compare to the withheld known preferences. In the dominant practice, this comparison has been oriented to measure the accuracy of rating prediction, computing error-based metrics. However, in terms of the effective utility of recommendations for users, there is an increasing realisation that the quality (precision) of the ranking of recommended items can be more important than the accuracy (error) in predicting specific rating values. As stated in [33], the research community has moved from the annotation in context task (i.e., predicting ratings) to the find good items task (i.e., providing users with a ranked list of recommended items), which better corresponds to realistic settings in working applications where recommender systems are deployed. As a result, precision-oriented metrics are being increasingly considered in the field. Yet there is considerable divergence in the way such metrics are applied by different authors, as a consequence of which the results reported in different studies are difficult to put in context and be compared.

3.2 Evaluation metrics

The evaluation of recommender systems should take into account the goal of the system itself. For example, in [34] the authors identify two main user tasks: *annotation in context* and *find good items*. In these tasks the users only care about errors in the item rank order provided by the system, not the predicted rating value itself. Based on this consideration, researchers have started to use precision-based metrics to evaluate recommendations, although most works also still report error-based metrics for comparison with state of the art

approaches. Moreover, one should also consider alternative performance criteria, like the novelty of the suggested items and the item coverage of a recommendation method. We describe the above types of evaluation metrics in the subsequent sections.

3.3 Error-based metrics

The evaluation of quality of recommendation algorithm is directly proportional to accuracy it offers and on the coverage. Accuracy is fractional measure of exact recommendation out of total recommendations, whereas coverage is fractional measure of items/object in the search space which the system is able to provide recommendation for. Accuracy metrics are divided into statistical and decision support accuracy metrics. Statistical accuracy metric is produced with the difference between actual and predicted ratings of users. Traditionally, the most popular metrics to measure the accuracy of a recommender system have been the **Mean Absolute Error** (MAE), and the **Root Mean Squared Error** (RMSE): The measurement of deviation of recommendations from users specific value is termed as mean absolute error (MAE).The formula for MAE is as follows:-

$$\mathbf{MAE} = \frac{1}{N}\sum_{u,i}|p_{u,i} - r_{u,i}| \quad \text{.....[e]}$$

Where,

$p_{u,i}$ = Predicted rating for user u on item i,

$r_{u,i}$ = Actual rating

N = Total number of ratings on the item set.

Less the value of MAE, better is the recommendation engine. Whearas, The Root Mean Square Error (RMSE) formula is:-

$$\mathbf{RMSE} = \sqrt{\frac{1}{n}\sum_{u,i}(p_{u,i} - r_{u,i})^2} \quad \text{.....[f]}$$

RMSE emphasized on absolute error. For better recommendation accuracy, RMSE should be lower.

A critical limitation of these metrics is that they do not make any distinction between the errors made on the top items predicted by a system, and the errors made for the rest of the items. Furthermore, they can only be applied when the recommender predicts a score in the allowed range of rating values. Because of that, log-based, and some content-based and probabilistic recommenders cannot be evaluated in this way, since they depend upon a probability or, in general, a preference score. Hence, these methods can only be evaluated by measuring the performance of the generated ranking using precision-based metrics.

3.4 Precision-based metrics

- ***nDCG***

It is a fact that recommender suggests some items based on some relevance score, basically a non-negative number, called a gain in nDCG. In case, no user feedback is available, gain is set to zero. By adding all gain, we get cumulative gain. It is used to put most exact items on top. So, before calculating these courses every individual score is divided by increasing number (generally log of position of item) known as discounting (DCG). But, for using DCG's for comparison between users, we need to stabilize them.

- **Precision**

It can be observed as percentage of recommendations that are measured good. More the precision, better the recommendations. It ranges from 0 to 1. Formulae is:

Precision = Correctly recommended items / Total recommended items

- **Recall**

It can be observed as the percentage of good recommendations that are also part of recommended items. Lesser the recall value, better are the recommendations. It's formulae is :

Recall = Correctly recommended items / Total useful recommended items

- **Fallout/F-measure**

F-measure deals with both precision and recall. The formula of F-measure is given by:-

$$F\text{-measure} = \frac{2PR}{P+R} \ldots\ldots\ldots [g]$$

Where P stands for Precision and R stands for Recall..

3.5 Experimental setup

An important decision in the experimental configuration of a recommender evaluation is the dataset partition strategy. How the datasets are partitioned into training and test sets may have a considerable impact on the final performance results, and may cause some recommenders to obtain better or worse results depending on how this partition is configured. Although an exhaustive analysis of the different possibilities to choose the ratings/items to be hidden is out of the scope of this thesis, we briefly discuss now some of the most well-known methods used.

First, we have to choose whether or not to take time into account [35]. Time-based approaches naturally require the availability of user interaction data timestamps. A simple approach is to select a time point in the available interaction data timeline to separate training data (all interaction records prior to that point) and test data (dated after the split time point). The split point can be set so as to, for instance, have a desired training/test ratio in the experiment. The ratio can be global, with a single common split point for all users, or user-specific, to ensure the same ratio per user. Time-based approaches have the advantage of more realistically matching working application scenarios, where "future" user likes (which would translate to positive response to recommendations by the system) are to be predicted based on past evidence. As an example, the well-known Netflix prize provided a dataset where the test set for each user consisted on her most recent ratings [36].If we ignore time, there are at least the following three strategies to select the items to hide from each user: a) sample a fixed number (different) for each user; b) sample a fixed (but the same for all) number for each user, also known as *given n* or *all but n* protocols; c) sample a percentage of all the interactions using cross-validation. The most usual protocol is the last one [37], although several authors have also used the all butnprotocol [38].

Furthermore, as mentioned earlier, in order to evaluate ranked recommendations for a target user, it is required to select two sets of items, namely relevant and not relevant. In the next chapter, we describe different possibilities explored in the literature, along with a detailed analysis of these alternatives and the possible biases that may appear.

3.6 Summary

In the Recommender Systems literature several evaluation metrics, protocols, and methodologies have been defined. It remains unclear the equivalence between them and the extent to which they would provide comparable results.

The problem of evaluating recommender systems has been a major object of study and methodological research in the field since its earliest days. Error-based metrics have widely dominated the field, and precision-based metrics have started to be adopted more recently. Other metrics from the Machine Learning field have been proposed but they are not widely used in the community yet. Moreover, metrics for additional dimensions such as novelty or diversity have also started to be researched in the last few years. There are, still, important characteristics of the evaluation methodologies and metrics that remain unexplored.

Chapter – 4

Design & implementation of RWYW mobile recommender using Tanimoto, Euclidien& Pearson correlation similarities

4.1 Introduction

Development of big data analytics has been supported and allowed by software framework known as Hadoop.It is capable of developing various and intense applications of commodity hardware having large clusters. It also gives the scalability features to the user as they can add nodes, if requires.Base of Hadoop isMapReduce: distributed computing paradigm. It is based on map and reduce,user defined functions,specified on ajob. The functioning of a job can be seen as divider of input dataset into different independent blocks, parallel processed by map tasks.HDFS: Hadoop Distributed File System is responsible to create various replications of data blocks and to assign them to computer nodes across a cluster for more reliable communications computations. On the other hand, Mahoutis a machine learning library that runs on a Hadoop system. To solve problems like classification, clustering, and prediction,it has ranges of variousalgorithms. It provides solutions to intense and critical machine learning problems with the help of MapReduce Paradigm conjunction with Hadoop. Although Mahout is in developing stage, majorly contributes in construction of recommendation engines, typical web problems, and some document classifier. Both Mahout and Hadoop are free of cost and open source projects, their key characteristics support them to be good platform to provide solution to data intensive problem [39]. Recommender systems arepredictive system which filter the information for finding rating or preference, user would most likely assign to an item. Many researchers create recommendation engine using Mahout/Hadoop to generate fruitful results. [40] Proposedmobile phone recommendation system based on phone interest. They had used Phone interest as a latent level concept which is extracted from a group of r's webuselog data, who have the similar type of mobile phone. Theauthor then created a novel probabilistic model named "Phone Interest Model" [41] presents SMARTMUSEUM, a mobile ubiquitous recommender system for the Web of Data, and its application to information needs of tourists in context-aware, on-site access to cultural heritage. It utilizes Semantic Web languages as the form of data representation.[42] introduced recommender SMoCoR: smart mobile contact recommender, taking data of smart phone as an input. Itsuggests a way to contact friends, keeping in mind the current condition of a person. Itpasses the messages between friends, tell user whether it

is the right time to call friends and also recommends the most suitable way for contacting friends.To the best of our knowledge, the work we presented is unique and brand new idea to facilitateboth consumer and cellular companies by providing a recommendation system which exactly provides what they want based on gender, locality, and plan subscriptions. Recently, mobile marketing has been recognized as a leading field that explores and use big data analytics to meet customer expectations, to be in competitive environment successfully and to bear economic pressure gracefully.

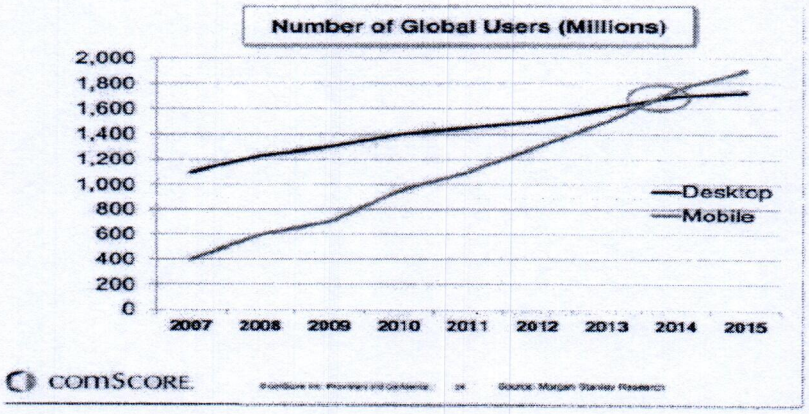


Figure4.1: Yearly increase in mobile devices [Available: www.comscore.com/layout/set/popup/.../2/.../2015_US_Mobile_App_Report.pdf]

When it comes to mobile users, the following are the various aspects of user for using his cell phone:-

1. Surf digital media

2. Calling

3. Messaging

4. Exploring basic features [camera, calculator, to do list etc.]

As a network providing company, the main concern is how to increase business and network while providing best services to the user. For all these services users have to pay. There are the tariff, SMS, data and roaming plans which user must subscribe to meet his needs. Companies generally inform their users about such plans through SMS. But the crude fact is, it's very annoying for the customers to receive all these messages which are of no use for them. Today an individual is so occupied by innumerable tasks and timelines that he feels

that involving him in all this is just a waste of time, How about a recommendation system that hits his interest specifically and stops all other irritating messages.Every competitive company wants to attract new customers, while retaining the exciting ones. In short, they want to churn out customers at any cost. This concept become more important in today's environment where customers have a numerous opportunities for fulfilling their needs. All they want is best services in minimum cost. And if they get it from another provider, they change their network without giving a second thought. In such cases, providing plans that suits to customer interest is a great idea.

4.2 Proposed Approach

In the proposed approach, the item is taken as current running airtel plans such as tariff, SMS, data, and roaming. A user profile includes information about ratings of their used plans along with their basic details. The batch system includes statistical analysis of the dataset, definition of a number of models and similarities and accuracy evaluation for user's profiles. Recommendation module is responsible for the selection of best model and for producing the recommendations. Finally, recommendations as an output are passed to consumer mobile devices.

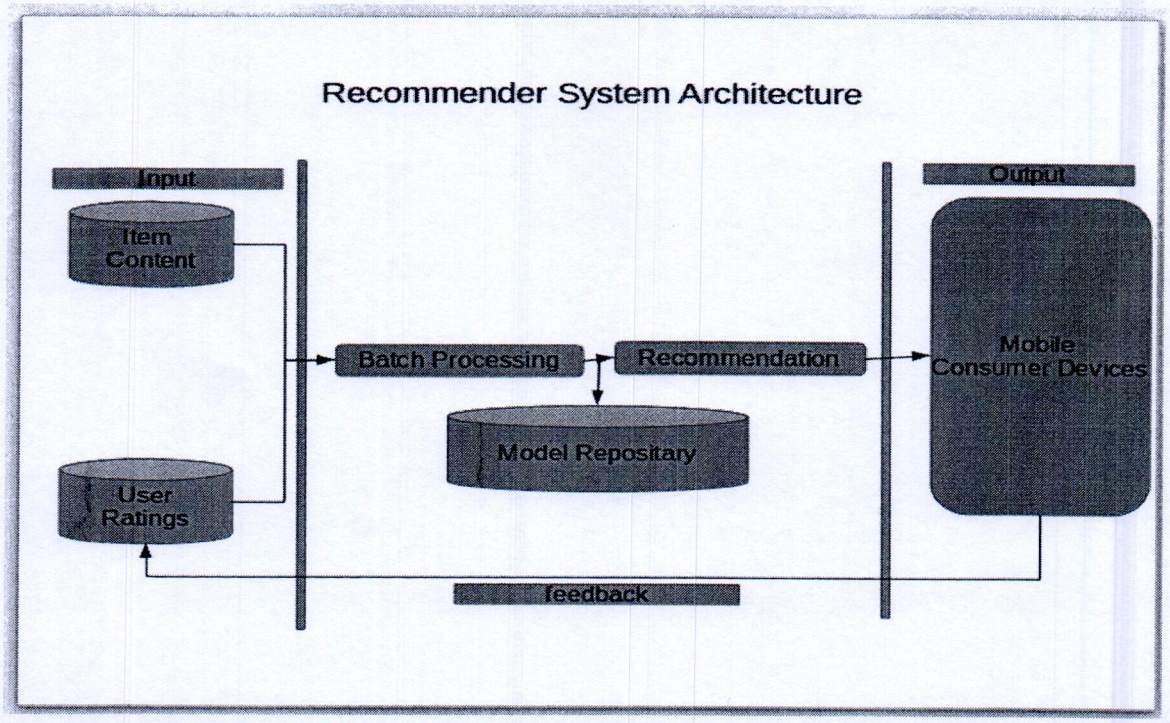


Figure4.2: Proposed RWYW Architecture

4.2.1 Input

Input is data on which recommender will base its recommendations. Generally, in Mahout, this data is in the form of preferences. We can see preferences as associations from users to an item, where users and items could be anything. A preference consists of User ID, Item ID some form of strength measure expressing the user preference for the item. In our experiment, we have used comma separated value format as an input. It is derived from Item Content and User profile.

4.2.2Item Content

In our experiment, the item content consists of all cellular plan of Airtel of a single locality. The total plan chosen is 203 based on talk time, data, roaming, SMS. The details of all these plans are available on [43]. For the understanding of the item, some of the item chosen with their item id is as follows:-

Category	Airtel plan Details					Item ID
Data Packs	2G	3G	4G	Other benefits	Days	
	90MB				3	10
		250 MB			28	17
		1.25GB		Wynk plus internet	28	25
Mixed	Data	STD	Local	Landline	SMS	
	110 MB 2G	2.5p/s	2p/s	2.5p/s		68
	7GB 3G	Unlimited	unlimited	unlimited		75
						1Q
FC	Talktime	Incoming	Outgoing	Minutes		
				85		92

	1479	20p/s	30p/s			97
SMS	Local	National	International	Days		
	150	175		28		100
	300	300		28		110
Tarriff	Minute'slocal	Minute'slocal	Local call	STD call	Talktime	
					1050	159
	90					170
			40p/s	40p/s		200

Table 4.1 Airtel mobile plans

This data is then processed to get valuable information and convert it into recommender's input format.

4.2.3 User Profile

It consists of user information such as User ID, data plan they used, preferred timings to use data packs, types of calls they make generally local or STD, age, sex, their locality and the ratings for plans they have used. For our experiment, we have taken data from users of the same locality via using the online form. This data is then processed to get useful information and to convert it into recommender's input format.

4.2.4 Batch Processing and Model Repository

The batch system consists of statistical analysis of the dataset. The data came from user profile and item content is analyzed as a dataset on which recommender will recommend its recommendations. Then this will be passed to the recommendation engine. The model repository consists of description of a number of models and likeness and accuracy evaluation for user's profiles. Some similarity models include in this are as follows:-

- **Pearson correlation-based Similarity**
- **Cosine measure Similarity**
- **Euclidean distance Similarity**
- **Tanimoto coefficient Similarity**

4.3 Recommendation and output

Recommendation module is accountable for the selection of finest model; apply it on the data set, divide the data set into test and training data and to generate recommendations. The algorithm used in the implementation framework is directly proportional to recommendation quality. Two recommenders are different from each other on certain basis:-

- Analysis of data
- Similarity measure used
- How relevant preference has been calculated

The work of recommendation module is to store the result in the form of preferences and it also encapsulates recommender input data in Mahout.

4.4 Algorithm

User-based recommender algorithm is the process of suggesting items to some user. In the algorithm below, some users are denoted by 'u'. The algorithm is as follows:- [44]

for every item i that u has no preference for yet

for every other user v, that has a preference for i

compute a similarity's between u and v

incorporate v's preference for i, weighted by s, into a running average

return the top items, ranked by weighted average

In our experiment, we ran it on three similarities i.e. Tanimoto, Euclidean Distance, and Pearson Correlation. Then these recommendations can be passed to user mobiles for their best-suited plans according to their gender and interest like women interest in data packs is very different than men. The feedback can be used to improve the performance of the recommender.

4.5 Results and Discussions

Most recommender engines operate by trying to calculate item's expected ratings. Experiments are performed to evaluate, how good is the performance of RWYW for offering correct plans to the user. We have taken 50000 users generating 1.5 Lakhs of data, among

these users 25000 are males whereas 25000 are females. For checking the performance of the approach, we have to conduct various experiments, so that the outcomes of these experiments provide us the clear picture of drawbacks of our recommendation engine. Further, these outcomes provide a base for future improvised recommender. Machine learning is important factors of recommendation, two phases which are included in it are: training and testing. . It is critically important to train the model before performing analysis on new data sets. For our experiment, we have taken 80% of data as training data set and 20 % of data as testing data set. Difference in the interest and basic temperament of men Verses women give rise to the difference in their mobile usage. For example, the women are more talkative and much more interested in online shopping in contradiction with men who are more interested in small talks and games that means men generally prefer lesser amount talk time packs and high GB internet packs with high speed mostly 3G/4G data plans whereas women need special talk time plans like free minutes plans and low GB internet with just fair speed. That's why we had divided are data sets in two parts men and women because there is much more similarity in choice of mobile packs of a same gender. Another parameter of our data set is the region, user belongs. Similar region add on the chances of habit similarity between two users of same place. Although, key parameter is user ratings which they have given to used packs. All this data is passed to the Generic User Based recommender, where first it is checked whether the user is male or female, according to which it picks the data set, then forms a similarity metrics of users based on the packs they are using, belong to which area, which objects are similar to their taste, calculates a similarity measure of it and finally suggest the plans which have highest similarity measure for the particular user. i.e. there are very bright chance that user appreciates the plan and starts using it. Our work shows that RWYW is quite easy to use and can be useful to improve the quality of services.

4.6 Evaluation

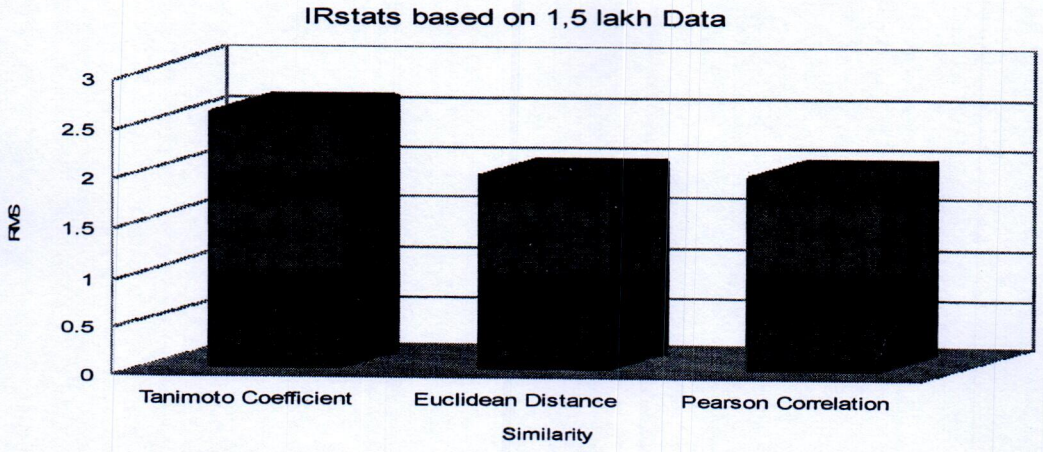
The evaluation of quality of recommendation algorithm is directly proportional to accuracy it offers and on the coverage. Accuracy is fractional measure of exact recommendation out of total recommendations, whereas coverage is fractional measure of items/object in the search space which the system is able to provide recommendation for. Accuracy metrics are divided into statistical and decision support accuracy metrics. Statistical accuracy metric is produced with the difference between actual and predicted ratings of users.

- **RMSE**

into statistical and decision support accuracy metrics. Statistical accuracy metric is produced with the difference between actual and predicted ratings of users.

- **RMSE**

RMSE emphasized on absolute error. For better recommendation accuracy, RMSE should be lower. For our experiment, we have taken ratings between 1 to 5 so, RMS gives the difference between actual and estimated value in testing set example. In our experiment, for Tanimoto Coefficient it is 2.5, which is not good in scale of 1 to 5. Figure 4.3 shows the RMS error measure with respect to similarity measure concludes that the suggested recommendations are better with Euclidean distance similarity and worst with Tanimoto Coefficient similarity.



Graph 4.3: RMS error Verses similarity

Decision support accuracy metrics

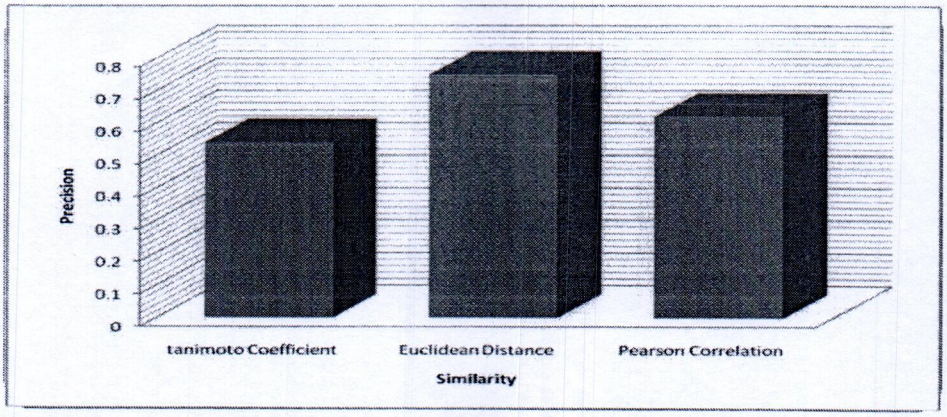
- **Precision**

Graph 4.4 demonstrates the precision measure with respect to similarity measures .It concludes that suggested proposals are improved with Euclidean distance. In contrast, with Tanimoto similarity only 54% suggestions are of any importance to users.

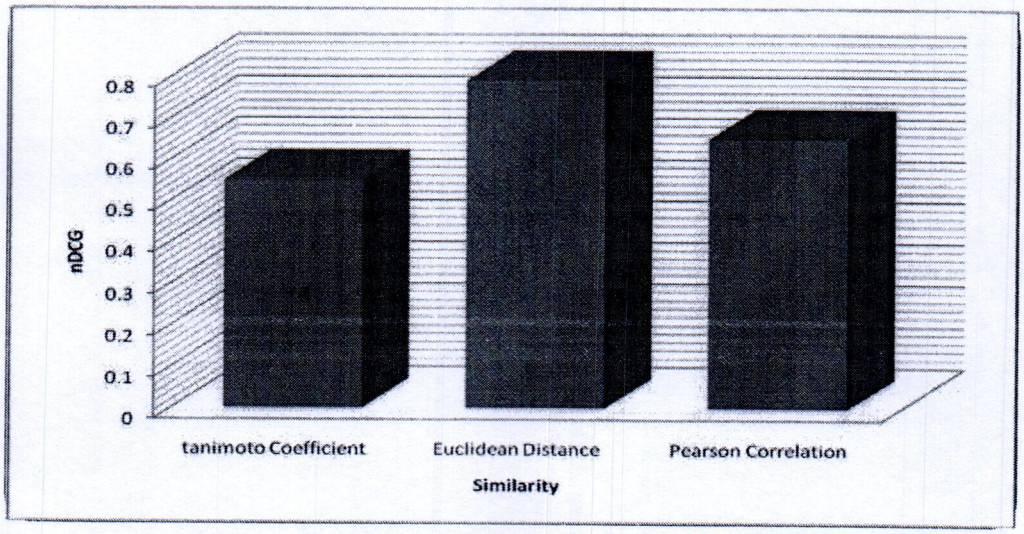
- **nDCG**

Graph 4.5 represents the result of nDCGVs similarity, greater the nDCG, better are the recommendations, it ranges between 0 to 1.

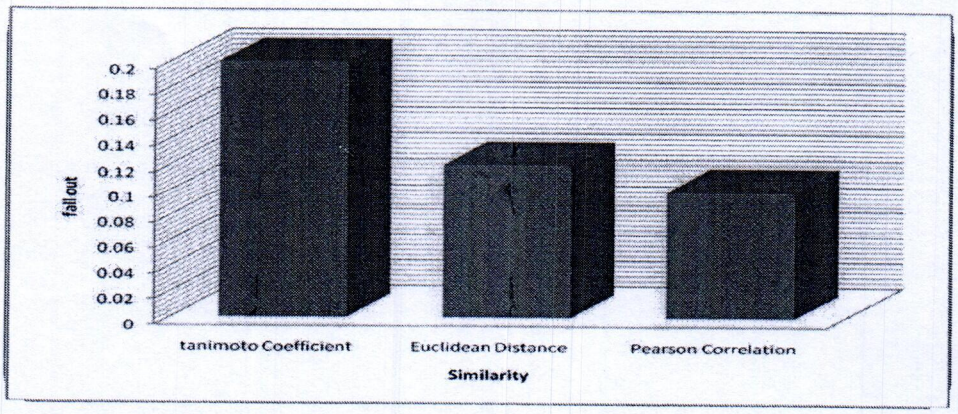
recommendation, that is not at all possible. Here, for Tanimoto Coefficient Similarity recommender fails 20 times out of hundred to provide good recommendation. Hence, we can conclude that the Euclidean Distance outperformed other similarity models.



Graph 4.4: precision Verses similarity



Graph 4.5: nDCG Verses similarity



Graph 4.6:SimilarityVs Fallout

At last, Table 4.2 gives the clear picture of the proposed recommendation engine execution time with respect to neighbourhood size

S.NO	Similarity Measure	Size of dataset	Neighbourhood Size	Execution Time (sec)
1	Tanimoto Coefficient Similarity	1.5 lakh	200	420
			5	402
2	Euclidean Distance Similarity	1.5 lakh	200	120
			5	98
3	Pearson Correlation Similarity	1.5 lakh	200	140
			5	112
S.NO	**Similarity Measure**	**Size of dataset**	**Neighbourhood Size**	**Execution Time (sec)**
1	Tanimoto Coefficient Similarity	1.5 lakh	200	420
			5	402
2	Euclidean Distance Similarity	1.5 lakh	200	120
			5	98
3	Pearson Correlation Similarity	1.5 lakh	200	140
			5	112

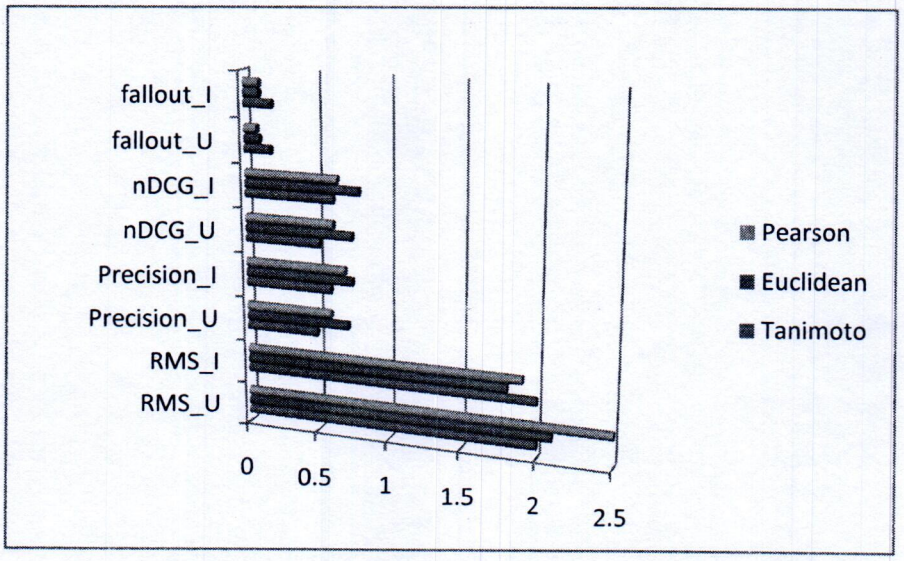
Table4.2: Neighbourhood size impact on execution time

4.6 Summary

In the current chapter, "RWYW" is presented which facilitates smooth user interaction between the user and cellular network providing companies. It saves the time of the user and

From all these experiments, which are implemented recommender algorithms through the Mahout java library (an open-source implementations of machine-learning algorithms. It is powering several portals e.g. Yahoo! Mail and AOL).We have observed that latency of Item Based model is very much dependent on the selection of similarity measure. The fastest recommendations are output of Euclidean and Pearson similarities. Tanimoto is the slowest and does not scale well with the density growth.

From the Graph 5.6, when the data is same that is users are same as well as items are same[in this case we have taken dataset 1i.e50,000 users generating 1.5 Lakhs data], we have found that the performance of item recommender is better than user based recommender[54].Hence, we concluded the item recommender is better than User based recommender, at least in suggesting right mobile plan to the user.



Graph 5.6: Comparison Graph of User based & Item based Recommender

5.4 Summary

In this chapter we have discussed about item based recommendation method which is capable of distributing right mobile plan to the right user. It is beneficial to both, the user of network and for the network providing company. We successfully implemented it and results show that among three methods of similarity calculation, Ecludien and Pearson perform better than Tanimoto in sorting accuracy and hitting rate.

Chapter – 6

Efficiency Booster For Recommendation System

6.1 Introduction

Recommender systems can be seen as guides, which helps users to find product of their interest .The exponential evolution of the world-wide-web along with the hype of e-commerce has led to the expansion of recommender system .It is a personalized information provider to recognize item sets that will be of interest to a particular user. Recommender systems are the base for future of the smart webs. The systems produce smooth user experience by making information retrieval easier and divert users from queries typing phase towards hit it off suggested links. No one is untouched by real-life recommender systems .They are doing amazing work, when browsing for music, movies, news or books. These engines are mandatory for websites like Amazon, Myntra or Netflix. In fact the core of recommender system is its algorithm; it has been characterized into four methods, namely graph model recommendation, collaborative filtering recommendation, hybrid recommendation and content-based recommendation [56]. On the basis of different approaches used for development of recommender systems such as demographic, content, or historical information [57,58,59], User based collaborative filtering came out as the most popular and promising for building recommender systems to date .Yet it is most successful but unluckily, the linear growth of its computational complexity with the count of customers can grow to be several millions as witnessed in commercial applications. To overcome this scalability issues, the item-based recommendation techniques have been developed .It generates user-item matrix to recognize relations among the diverse items, and use them to produce recommendations [60,62]. The progress of RS has shown the significance of hybrid techniques, which is the combination of different techniques in order to utilize the benefits of each of them. A survey on hybrid recommenders has been discussed in [63]. While the current surveys focus on the most apt methods and algorithms of the recommendation system, our paper instead tries to focus on how to increase the efficiency of the RS: from a traditional phase to modified one.

In the current chapter, we present an mobile plans item recommender using similarity fusion measure.

6.2 Item recommender Architecture using Similarity fusion

For the item recommender using similarity fusion, the architecture is as shown in Figure 6.1. It consists of five stages. Let's discuss all these stages in detail:

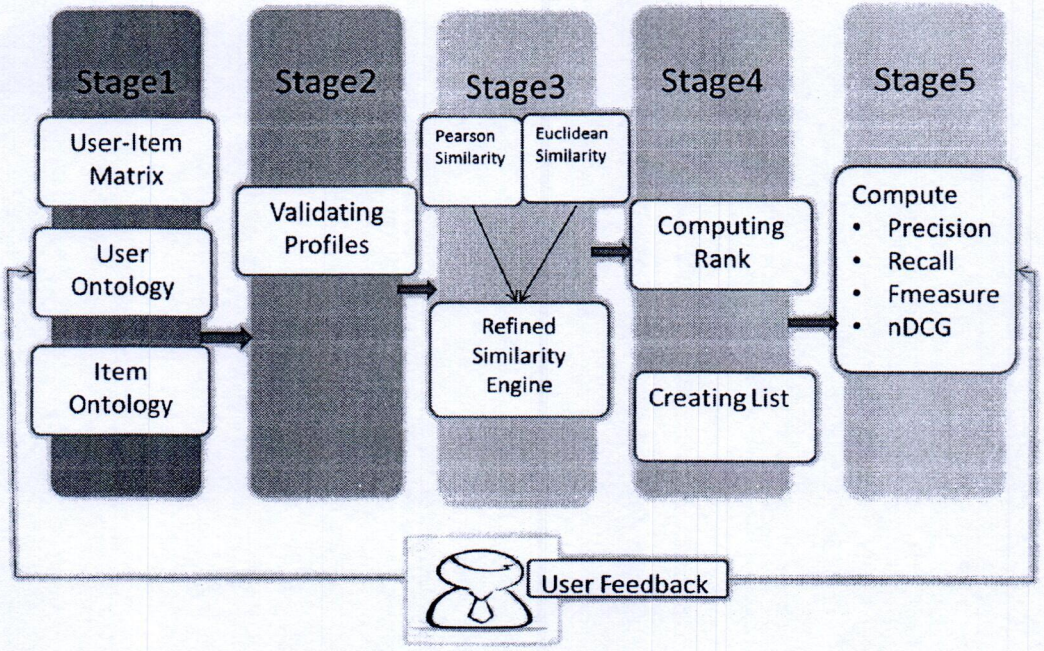


Figure 6.1: Architecture of mobile plans item recommender using similarity fusion

- **Stage 1:Data Creation**

In Stage 1, there are three main tasks. First is the creation of user-item matrix. In our experiment, we have considered the user experience of using particular item. User can rate them in between 1-5. Here item means particular mobile plans like tariff, ,top up, internet, SMS, wynk etc. So on the basis of these ratings user-item matrix is formed. Next is user ontology, which is based on mobile users information like sex, age, locality, occupation, Income and usage of basic facilities of mobile such as calling, net surfing and texting. And the last is item ontology, as we have considered Airtel as network provider for mobiles ,hence in our experiments item is the mobile plans provided by Airtel network for an example 220 Rs full talk time is one of the item for our recommender . Further, Item ontology contains item id, item description, item raking given by user. All of these data is then passed to second stage.

- **Stage 2: Profile Validation**

After completion of the stage 1, it is clear that which data is valuable and useful i.e which data should be discarded and which is of use. So fruitful data is then processed to be passed

to recommender. This is very important for making good recommendations and here this stage is termed as profile validation. Its output is then passed to Stage 3.

- **Phase 3:Similarity Fusion**

The similarity has been calculated based on the items that user with similar liking opted previously and on the basis of this similarity measure, the recommendations to the active users will be given. Traditionally, any one of the similarity model namely : Pearson correlation similarity, Euclidean distance similarity, Spearman correlation similarity, Cosine measure similarity and Tanimoto coefficient similarity is used. But in our experiment, we used fusion similarity measure. It is the combination of the two similarity measure namely pearson and Euclidean. Pearson is a range of numbers between -1 to +1. According to this similarity, the two series share the linear relationship, when we measure the tendency of two numbers move proportionally. Whereas, in Euclidean distance users are seen as points in multidimensional space and the coordinates of these points are seen as preferences value. The Euclidean similarity measure is inversely proportional to the Euclidean distance. But in fusion method, we calculate the similarity of the common item by adding both similarity measures that in turn increases the similarity measure of item , hence help recommender to place most apt item plan to the customer. It clearly effects the accuracy of recommender.

Algorithm is as follow:

For every item I1 in item ontology

For every user U in user ontology who bought I1

For each item I2 bought by

user U

Record that a user bought both I1

and I2

For each item I2

Calculate the similarityS1 between I1 and I2 using Pearson

Calculate the similarityS2 between I1 and I2 using Euclidean

Calculate the Improvised similarity between I1 and I2 by adding S1and S2

- **Stage 4: Computing rank and creating list**

After the completion of stage 3, now we need to calculate the' n' neighbour of the current user 'u' and pass neighbourhood size as a parameter in mobile plan Item based Recommender. One of the K-nearest neighbour algorithms is used for the recommendation process. Size of the neighbourhood is important for the accuracy of the algorithm. The utmost similar users are taken (the number is limited by the neighbourhood size). A list of most suitable items for recommendation has been generated and passed to user.

- **Stage 5:Evaluation**

How good is our recommender, it depends upon performance evaluation. This can be judged by parameters like recall, F-measure, precision and nDCG. In this stage all of these parameters are calculated. The detailed description of these parameters is covered in next section.

6.3 Experimental Results

To study the effect of similarity fusion on recommender, various experiments had been performed with the help of datasets. The different datasets used in our experiments are as follows:

I. 1.5 lakhs of data generated by 50 thousands of mobile users

II. 2 Lakhs of data generated by 70 thousands of mobile users

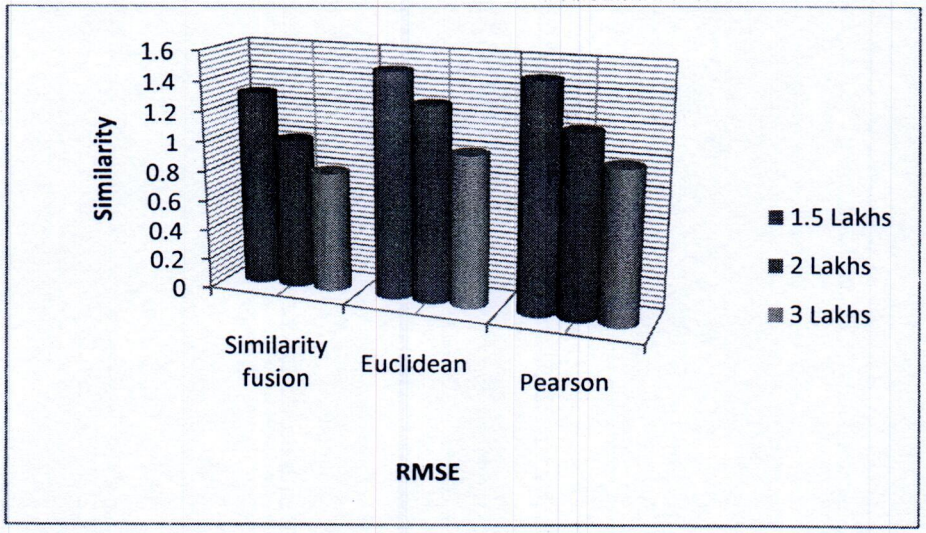
III. 3 Lakhs of data generated by 1 lakh of mobile users

For applying machine learning, for our experiment, we have divided the data set into two parts i.e for testing and training. We have chosen the ratio of 80:20 for training testing phase .Various parameters are considered for these experiments such as Income, sex, agearea ,generic interest such as calling, texting , surfing, or all. Finally, all this data is transferred to the mobile plan Item Based recommender, where data processing takes place and then generates the similarity metrics to create the list of most appropriate items for the particular users according to their need. And with the help of similarity fusion, there is quite bright chance that user appreciates the plan and starts using it.

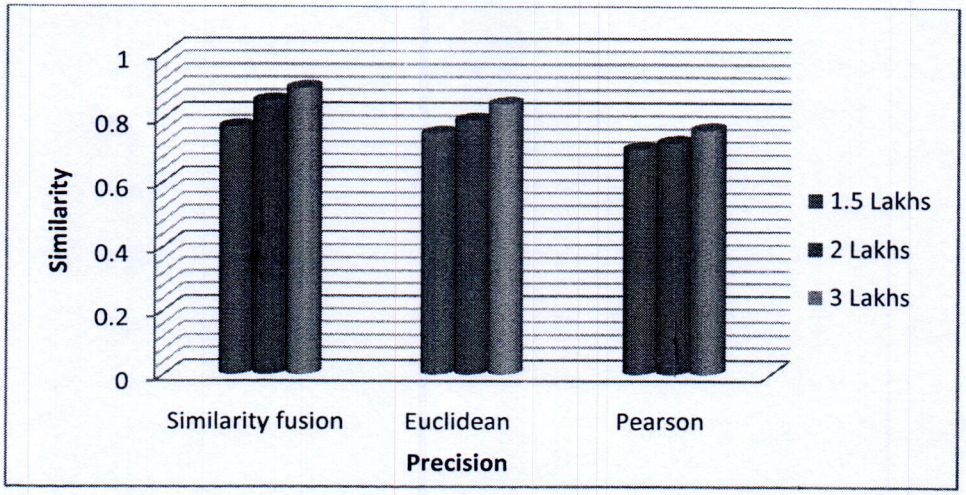
6.4 PerformanceAnalysis

For checking the accuracy of the designed recommender, we need to evaluate the measuring criteria such as Mean absolute error (MAE),Recall ,Precision, fall out and nDCG:

For our experiments, we use mahout and its library[55] .Graphs 6.2, 6.3, 6.4, 6.5 is the representation of the evaluation results of our item based recommender using similarity fusion on the considered data sets. Results show that the fusion method used in this proposed approach is a enhancer for suggesting the best suited mobile plans to the particular customer .Further, We have observed that latency of Item Based model is very much dependent on the selection of similarity measure. The fastest recommendations are output of Euclidean and Pearson similarities. But similarity fusion outstands both in terms of accuracy.



Graph 6.2: Similarity fusion effect on RMSE



Graph 6.3: Similarity fusion effect on Precision

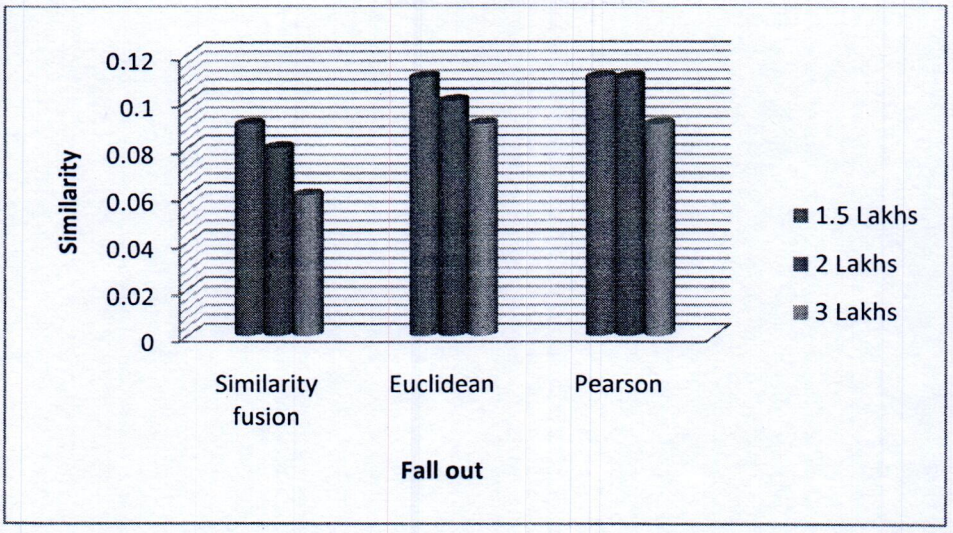
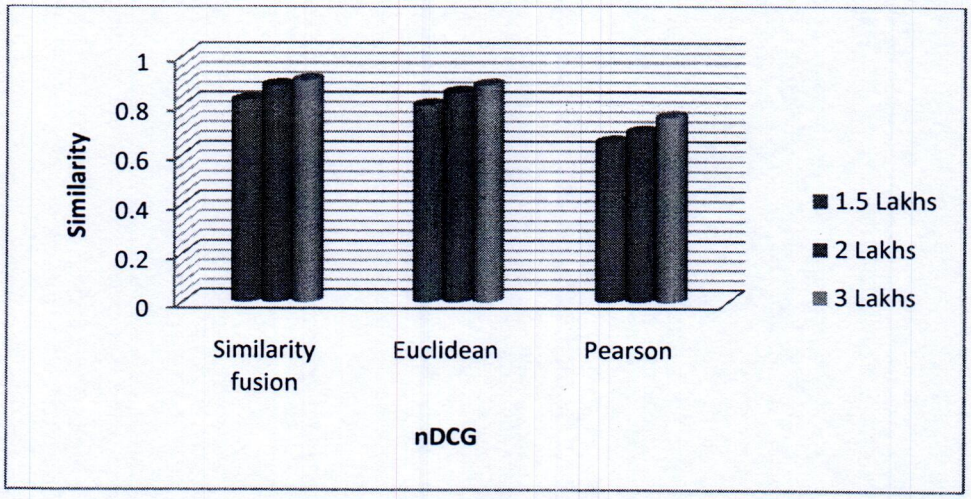


Figure 6.4: Similarity fusion effect on Fall out



Graph 6.5: Similarity fusion effect on nDCG.

6.5 Summary

In this chapter we proposed a mobile plans recommendation method using similarity fusion which is capable of distributing more accurate mobile plan to the appropriate user. Hence, it is capable of enhancing user satisfaction level towards its chosen network. We effectively implemented it and results show that the similarity fusion method clearly enhances the accuracy and hitting rate, compared to individual use of Pearson and Euclidean similarity method in recommendation process. Although, experiment results show that the proposed similarity fusion method provide better recommendation quality but increase in time is basic factor to control.

Conclusion & Future work

RS is quite useful and a driving force for today's business. Its utility in online shopping is remarkable.With enhancement in recommendation approaches, the efficiency of RSs has also increased. In this thesis we have proposed a mobile plans recommender &investigated how to measure and predict the performance of recommender systems.

Moreover, since we aimed to predict the performance of a particular recommender system, we required a clear recommender evaluation methodology against which performance predictions can be contrasted. Hence, in this thesis we addressed the evaluation methodology as part of the problem,

Through a series of empirical experiments on several datasets and experimental designs, we have found a correspondence between the predictive power of our performance predictors and performance enhancements in the mobile plans as an application.

We successfully implemented RYRW mobile plans recommender and experiment shows that Euclidean Distance is the best similarity model for this type of recommender system as minimum is neighbourhood size, the maximum is the performance

Moreover, a novel item based recommendation method is also produced, which is capable of distributing right mobile plan to the right user. It is beneficial to both, the user of network and for the network providing company. We successfully implemented it and results show that among three methods of similarity calculation, Ecludien and Pearson perform better than Tanimoto in sorting accuracy and hitting rate.

With time CF has become a striking subject for researchers because of its competency to handle information overload efficiently. Many techniques and different approaches have been developed to generate an effective recommendation. This is interesting as different application's scenarios could have a fittest solution. Considering this, finally we have discussed the different techniques that are capable to enhance the efficiency of RS and proposed similarity fusion technique to provide better prediction accuracy. We plan to extend the proposed algorithm in a real-world recommender system and evaluate its efficiency.

Result shows similarity fusion method is capable of distributing more accurate mobile plan to the appropriate user. Hence, it is capable of enhancing user satisfaction level towards its chosen network. We effectively implemented it and results show that the similarity fusion method clearly enhances the accuracy and hitting rate, compared to individual use of Pearson and Euclidean similarity method in recommendation process. Although, experiment results show that the proposed similarity fusion method provide better recommendation quality but increase in time is basic factor to control.

In future, we plan to experiment on distributed data; we try to fit this recommendation engine on various other applications such as online banking loans, student high education in which it will recommend promising college list for students want to go for higher education.

Further, wehave planned to evaluate how the performance of recommender systems is altered when they are implemented in a distributed environment and how the recommender efficiency can be increased.

LIST OF PUBLICATION

(A) International Journals:

Sr. No.	Name of the Journal	Title of the paper	Published in
1.	International Journal of Computing and Engineering Research(IJCER)	*Efficiency Booster Techniques For Recommendation System*	**Accepted ,will be published in October**
2.	International Journal of Computer Science and Information Security (IJCSIS) ISSN: 1947-5500 **(SCOPUS Indexed)**	*A Novel Item Recommender for Mobile Plans*	Vol. 16 No. 9 ,pp. 58-62, September 2018
3.	Journal of Mechanics of Continua and Mathematical Sciences(JMCMS) ISSN : 0973 – 8975 **(E-SCI ThomsonsReauters Indexing)**	*An Improvised Recommendation System for Mobile Plans Using Similarity Fusion*	Vol. 13 No. 4 , September – October 2018,Will be published in Oct End
4.	International Journal of Innovations & Advancement in Computer Science. IJIACS, ISSN 2347–8616. **(UGC Approved Journal)**	*RWYW: Recommend What you Want-A Recommender for Mobile Plans*	Volume 7, Issue 2, February 2018
5.	International Journal of Innovations & Advancement in Computer Science. IJIACS, ISSN 2347–8616. **(UGC Approved Journal)**	*Recommendation System: Boon to Consumer and Producer*	Volume 7, Issue 2, February 2018

(B) International Conferences:

Sr. No.	Name of the conference	Title of the paper	Venue
1.	3rd International Conference on Computing: Communication, Networks and Security	*"RWYW: Recommend What you Want, A Recommender for Mobile Plans."*	Mody University, Lakshmangarh

2.	3rd International Conference on Computing: Communication, Networks and Security	*"Recommendation System: Boon to Consumer and Producer"*	Mody University, Lakshmangarh

(C) Certificate Courses

1. Online Course onMachine Learning with Mahout
https://www.edureka.co/mycertificate/211ab6812efb9c9315a557d17bfce845?utm_source=linkedin&utm_medium=cert&utm_campaign=linkedincert

Mapping of Objectives

S.No	Objectives	Paper(s)
1	To design and implement user based mobile recommender based on Tanimoto, Euclidean distance and Pearson correlation similarities.	[2]
2	To design and implement item based mobile recommender based on similar tiesmeasures such as Tanimoto, Euclidean distance and Pearson correlationsimilarities.	[3]
3	To evaluate the quality of mobile recommender using standard available methods.	[1]
4	To design, develop and implement efficiency booster technique for mobile recommender.	[4],[5]

References

1 N. Landia and S.S. Anand, "Personalised Tag Recommendation", in Proceedings of the 2009 ACM Conference on Recommender Systems, pp. 83–36, 2009.

2 S.K. Lee, Y.H. Cho and S.H. Kim, "Collaborative Filtering with Ordinal Scale-based Implicit Ratings for Mobile Music Recommendations", Information Sciences, Vol. 180, No. 11, pp. 2142–2155, 2010.

3 D.H. Park, H.K. Kim, I. Y. Choi and J.K. Kim, "A Literature Review and Classification of Recommender Systems Research", Expert Systems with Applications,Vol. 39, No. 11, pp. 10059–10072, 2012.

4 W. Carrer-Neto, M.L. Hernandez-Alcaraz, R. Valencia-Garcı´ a and F. Garcı´a-Sanchez, "Social Knowledge-based Recommender System, Application to the Movies Domain", Expert Systems with Applications,Vol. 39, No. 12, pp. 10990–11000, 2012.

5 P. Winoto and T.Y. Tang, "The Role of User Mood in Movie Recommendations", Expert Systems with Applications,Vol. 37, No.8, pp. 6086–6092, 2010.

6 A. Nanolopoulus, D. Rafailidis, P. Symeonidis and Y. Manolopoulus, "Music Box: Personalizad Music Recommendation Based on Cubic Analysis of Social Tags", IEEE Transactions on Audio, Speech and Language Processing,Vol. 18 No. 2, pp. 407–412, 2010.

7 S. Tan, J. Bu, CH. Chen and X. He, "Using Rich Social Media Information for Music Recommendation via Hypergraph Model", ACM Transactions on Multimedia Computing, Communications, and Applications,Vol. 7, No.1, pp. 213-237, 2011.

8 A.B. Barragans-Martı´nez, E. Costa-Montenegro, J.C. Burguillo, M. Rey-Lopez, F.A. Mikic-Fonte and A. Peleteiro, "A hybrid Content-based and Item-based Collaborative Filtering Approach to Recommend TV Programs Enhanced with Singular Value Decomposition", Information. Sciences, Vol. 180, No.22, pp. 4290– 4311, 2010.

9 E.R. Nunez-Valdez, J.M. Cueva-Lovelle, O. Sanjuan-Martı´nez, V. Garcı ´ a-Dı´az, P. Ordonez and C.E. Montenegro-Marı´ n, "Implicit Feedback Techniques on Recommender Systems Applied to Electronic Books", Computers in Human Behavior, Vol. 28, No. 4, pp. 1186–1193, 2012.

10 R. Gonzalez-Crespo, O. Sanjuan-Martı´nez, J. Manuel-Cueva, B. Cristina-Pelayo, J.E. Labra-Gayo and P. Ordonez, "Recommendation System Based on User Interaction Data Applied to Intelligent Electronic Books", Computers in Human Behavior, Vol. 27, No. 4, pp. 1445–1449, 2011.

11 J. Serrano-Guerrero, E. Herrera-Viedma, J.A. Olivas, A. Cerezo and F.P. Romero, "A Google Wave-based Fuzzy Recommender System to Disseminate Information in University Digital Libraries 2.0.", Information Sciences, Vol. 181, No. 9, pp.1503–1516, 2011.

12 C. Porcel, J.M. Moreno and E. Herrera-Viedma, "A Multi-Disciplinar Recommender System to Advice Research Resources in University Digital Libraries", Expert Systems with Applications,Vol. 36, No. 10, pp. 12520–12528, 2009.

13 C. Porcel and E. Herrera-Viedma, "Dealing with Incomplete Information in a Fuzzy Linguistic Recommender System to Disseminate Information in University Digital Libraries", Knowledge-Based Systems, Vol. 23, No.1, pp. 32–39, 2010.

14 J. Bobadilla, F. Serradilla and A. Hernando, "Collaborative Filtering Adapted to Recommender Systems of E-learning", Knowledge Based Systems, Vol. 22,No. 4, pp. 261–265, 2009.

15 J.J. Castro-Sanchez, R. Miguel, D. Vallejo and L.M. Lopez-Lopez, "A Highly Adaptive Recommender System Based on Fuzzy Logic for B2C E-commerce Portals", Expert Systems with Applications, Vol. 38, No. 3, pp. 2441–2454, 2011.

16 E. Costa-Montenegro, A.B. Barragans-Martı ´ nez, M. Rey-Lopez, "Which App? ARecommender System of Applications in Markets: Implementation of the Service for Monitoring Users' Interaction", Expert Systems with Applications, Vol. 39, No. 10, pp. 9367–9375, 2012.

17 Digital Universe Study (on behalf of EMC Corporation), "Big Data, Bigger Digital Shadows and Biggest Growth in the Far East", [Online] Available http://idcdocserv.com/1414, 2012.

20 Cisco Internet Business Solutions Group (IBSG), "The Internet of Things: How the Next Evolution of the Internet is Changing Everything",[Online], Available: http://www.Cisco.com/web/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf, 2012.

21 ComScore, "Mobile App Report", [Online]. Available: www.comscore.com/layout/set/popup/2/2015_US_Mobile_App_Report.pdf, 2015.

22 IBM, "The Essential CIO", [Online], Available: http://www-935.ibm.com/services/uk /cio/ pdf/CIE03073-GBEN-01.pdf, 2011.

23 K. Mcnally, M.P. O'mahony, M. Coyle, P. Briggs and B. Smyth, "A Case Study of Collaboration and Reputation in Social Web Search", ACM Transactions on Intelligent Systems and Technology, Vol. 3, No. 1, pp. 30-37, 2011.

24 J. Bobadilla, F. Ortega, A. Hernando and A. Gutierrez, "Recommender Systems Survey",Knowledge-Based Systems,Elsvier, Vol. 46, pp. 109–132, 2013.

25 Chun- Wei Tsai, Chin-Feng Lai, Han-Chieh Chao and Athanasios V. Vasilakos, "Big data analytics: a survey", Journal of Big Data,Open Journal Springer, Vol. 2, No. 1, pp. 1-32, 2015.

26 BerkovskyS.,Kuflik T. and Ricci,F.:Mediationof User Models for Enhanced Personalization in Recommender Systems, UserModelingandUser-AdaptedInteraction,Vol. 18, No. 3, pp. 245–286, 2008.

27 Berkovsky S., Kuflik, T. and Ricci F.,"Cross-Representation Mediation of User Models, User Modeling and User-Adapted Interaction,Vol. 19, No. 2, pp. 35–63, 2009.

28 Schafer J. B., Frankowski D., Herlocker J. and Sen S.,"Collaborative Fltering Recommender Systems," In the Adaptive Web,Springer Berlin / Heidelberg, pp. 291–324. 2007.

29 Mahmood T. and Ricci, F., "Improving Recommender Systems with Adaptive Conversational Strategies, In: C. Cattuto, G. Ruffo, F. Menczer (eds.) Hypertext, ACM,pp. 73–82, 2009.

30 http://www.cs.carleton.edu/cs_comps/0607/recommend/recommender/itembased.htm, Accessed March,2016.

31 Guy Shani and AselaGunawardana.,"Evaluating Recommendation Systems", Evaluation Matrix, pp. 1-43, 2017.

32 In Ricci, F., Rokach, L., Shapira, B., and Kantor, P. B. Editors, "Recommender Systems Handbook", Springer US, Boston, Chapter 8, pp. 257–297, 2011.

33 Gunawardana, A. and Shani, G.,"A Survey of Accuracy Evaluation Metrics of Recommendation Tasks., J. Mach. Learn. Res., Vol. 10, pp. 2935–2962, 2009.

34 Herlocker J. L., Konstan J. A., Terveen L. G., and Riedl J. T.,"Evaluating Collaborative Filtering Recommender Systems", ACM Trans. Inf. Syst., Vol.22, No. 1, pp. 5–53, 2004.

35 Jahrer, M., Töscher, A., and Legenstein, R.,"Combining Predictions for Accurate Recommender Systems", In Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD, New York, NY, USA, ACM, pp. 693–702, 2010.

36 Bourke S., McCarthy K. and Smyth B. Power to the people,"Exploring Neighbourhood Formations in Social Recommender System", In Proceedings of the fifth ACM Conference on Recommender Systems, RecSys,New York, NY, USA. ACM, pp. 337–340, 2011.

37 Bennett, J. and Lanning, S.,"The Netflix Prize", In *Proceedings of the KDD Cup Workshop* New York. ACM, pp 3–6, 2007.

38 Roeder T., Gupta D.and Perkins C. Eigentaste,"A Constant Time Collaborative Filtering Algorithm", *Inf. Retr.*, Vol. 4, No. 2, pp. 133–151, 2001.

39 Wang, J., de Vries, A. P. and Reinders M. J. T.,"Unified Relevance Models for Rating Prediction in Collaborative Filtering. ACM Trans. Inf. Syst., Vol. 26, No. 3, pp. 1–42, 2008.

40 Ashish Gupta, "Learning Apache Mahout Classification", ISBN-13: 978-1783554959, 2015.

41 Bozhi Yuan, Bin Xu, Tonglee Chung, Kaiyan Shuai and Yongbin Liu, "Mobile Phone Recommendation Based on Phone Interest", B. Benatallah: Wise 2014, Part I, LNCS 8786, Springer International Publishing Switzerland,pp. 308–323, 2014.

42 TuukkaRuotsalob , KristerHaavh , Antony Stoyanovg , Sylvain Rochef , Elena Fanie, RominaDeliaid, EetuMakel, TomiKauppinena andEeroHyvonen, "SMARTMUSEUM: A Mobile Recommender System for the Web of Data", Journal of Web Semantics: Science, Services and Agents on the World Wide Web,2013.

43 XiweiZhuang, Yanchun Sun and Kui Wei," SMoCoR: A Smart Mobile Contact Recommender Based on Smart Phone Data", IEEE 38th Annual Computer Software and Applications Conference, ISBN: 978-1-4799-3575-8, pp. 634-635, 2014.

44 Sean Owen, Robin Anil, Ted Dunning and Ellen Friedman, "Mahout in Action", ISBN-13: 978-1935182689, 2011.

45 Apache Hadoop, [Online]. Available: http://hadoop.apache.org, 2015.

46 Jianguo Liu, Tao Zhou and Binhong Wang, "Research Process of Personalized Recommendation System", Progress in Natural Science, Vol. 19, No. 1, pp. 1-15, 2009.

47 Breese J. S., Heckerman D. and Kadie C.,"Empirical Analysis of Predictive Algorithms for Collaborative Filtering",UAI-98 Proceeding of the fourteenth Conference on Uncertainty in Artificial Intelligence, pp. 43-52, July 24-26, 1998.

48 Resnick, P., Iacovou, N., Sushak, M., Bergstrom, P. and Riedl, J.,"Grouplens: An Open Architecture for Collaborative Filtering of Netnews", SIGCSCW-1994.

49 Schafer J., Konstan J. and Riedl J.,"Recommender Systems in E-commerce. In Proceedings of ACM E-Commerce-1999.

50 M. Deshpande and G. Karypis,"Item-based Top-n Recommendation Algorithms", ACM Trans. Inf. Syst., Vol. 22, No. 1, pp. 143–177, 2004.

51 B. Sarwar, G. Karypis, J. Konstan and J. Riedl,"Item-based Collaborative Filtering Recommendation Algorithms", In Proc. of the WWW Conference, 2001

52 Bamshad Mobasher, Honghua Dai, Tao Luo, Miki Nakagawa and Jim Witshire,"Discovery of Aggregate Usage Profiles for Web Personalization", In Proceedings of the WebKDD Workshop, 2000.

53 Lyle H. Ungar and Dean P. Foster,"Clustering Methods for Collaborative Filtering", In Workshop on Recommendation Systems at the 15th National Conference on Artificial Intelligence, 1998.

54 Greg Linden, Brent Smith and Jeremy York, "Amazon.com Recommendations: Item-to-Item Collaborative Filtering," IEEE Internet Computing, doi:10.1109/MIC.2003.1167344, Vol. 7, No. 1, pp.76-80, Jan./Feb. 2003.

55 Ms. Neetu Singh, Dr. Puneet Kumar & Prof. Anil Kumar Dahiya, "RWYW: Recommend What You Want"-A Recommender for Mobile Plans, International Journal of Innovations & Advancement in Computer Science,Volume 7, February 2018

56 The Mahout Website." http://mahout.apache.org/.

57 Jianguo Liu, Tao Zhou and Binhong Wang," Research process of personalized recommendation system", Progress in Natural Science, Vol. 19 No. 1, pp. 1-5, 2009.

58 Breese J. S., Heckerman D. and Kadie C. "Empirical Analysis of Predictive Algorithms for Collaborative Filtering", UAI 1998.

59 Resnick P., Iacovou N., Sushak M., Bergstrom, P. and Riedl J., "Grouplens: "An Open Architecture for Collaborative Filtering of Netnews", SIGCSCW, 1994.

60 Schafer J., Konstan J. and Riedl J., "Recommender Systems in E-commerce", In Proceedings of ACM E-Commerce, 1999.

61 Bamshad Mobasher, Honghua Dai, Tao Luo, Miki Nakagawa and Jim Witshire, "Discovery of Aggregate Usage Profiles for Web Personalization". In Proceedings of the WebKDD Workshop, 2000.

62 Lyle H. Ungar and Dean P. Foster, "Clustering Methods for Collaborative Filtering in Workshop on Recommendation Systems", 15th National Conference on Artificial Intelligence, 1998.

63 R. Burke, "Hybrid recommender systems: survey and experiments", User Modeling and User-Adapted Interaction, Vol. 12, No. 4, pp. 331–370, 2002..

64 Riondato M, DeBrabant J A, Fonseca R, Upfal E,"PARMA: A Parallel Randomized Algorithm for Approximate Association Rules Mining in Mapreduce", In Proceedings of the ACM International Conference on Information and Knowledge Management, pp. 85–9, 2012.

65 Beckmann M, Ebecken NFF, de Lima BSLP, Costa MA,"A User Interface for Big Data with Rapidminer",RapidMiner World, Boston, MA, Tech. Rep., [Online]Available: http://www.slideshare.net/RapidMiner/a-user-interface-for-big-data-with-rapidminer-marcel-beckmann, pp. 13-52, 2014.

66 Z. Zheng, J. Zhu and M. R. Lyu, ''Service Generated Big Data and Big Data as a Service: An overview'', in Proc. IEEE Int. Congr. Big Data, pp. 403–410, 2013.

67 A. Bellogín, I. Cantador, F. Díez, P. Castells and E. Chavarriaga, "An Empirical Comparison of Social, Collaborative Filtering, and Hybrid Recommenders", ACM Trans. Intell. Syst. Technol., Vol. 4, No. 1, pp. 1–37, 2013.

68 W. Zeng, M. S. Shang, Q. M. Zhang, L. Lü, and T. Zhou, ''Can Dissimilar Users Contribute to Accuracy and Diversity of Personalized Recommendation?", Int. J. Modern Phys. C, Vol. 21, No. 10, pp. 1217–1227, 2010.

69 J. Wu, L. Chen, Y. Feng, Z. Zheng, M. C. Zhou and Z. Wu,''Predicting Quality of Service for Selection by Neighborhood-based Collaborative Filtering", IEEE Trans. Syst., Man, Cybern., Syst., Vol. 43, No. 2, pp. 428–439, 2013.

70 G. Adomavicius and J. Zhang, ''Stability of Recommendation Algorithms,'' ACM Trans. Inf. Syst., Vol. 30, No. 4, pp. 23:1–23:31, 2012.

71 X. Li and T. Murata, ''Using Multidimensional Clustering based Collaborative Filtering Approach Improving Recommendation Diversity,'' in Proc.

IEEE/WIC/ACM Int. Joint Conf. Web Intell. Intell. Agent Technol., pp. 169–174, 2012.

72 R. D. Simon, X. Tengke, and W. Shengrui, ''Combining collaborative filtering and clustering for implicit recommender system,'' in Proc. IEEE 27th Int. Conf. Adv. Inf.Netw. Appl., pp. 748–755, 2013.

73 Chen et al., "Big data challenge: A data management perspective", Frontiers Comput. Sci., Vol. 7, pp. 157–164, 2013.

74 Jahrer M, Toscher A and Legenstein R, "Combining Predictions for Accurate Recommender Systems", In: Proceedings of the ACM Conference on Knowledge Discovery in Databases (KDD), 2010.

75 Lathia N, Hailes S, Capra L andAmatriain X, "Temporal Diversity in Recommender Systems". In: Proceedings of the SIGIR, Geneva, Switzerland, July 19–23, 2010.

76 Li Y, Zhai CX, Chen Y, "Exploiting Rich User Information for One-Class Collaborative Filtering,KnowlInfSyst 38", pp: 277–301, 2014.

77 Bambini R, Cremonesi P and Turrin R, "A Recommender System for an IPTV Service Provider: A Real Large-Scale Production Environment", In: Ricci *et al* (eds) Recommender Systems Handbook. Springer, New York, NY, 2011.

78 Singh Reddy," A Survey on Platforms for Big Data Analytics". Journal of Big Data, Springer Open Journal, pp. 1-8, 2014.

79 Blondel VD, Decuyper A and Krings G,"A Survey of Results on Mobile Phone Datasets Analysis", EPJ Data Sci., pp. 4-10, 2015.

80 Miritello G, Rubén L, Cebrian M, Moro E," Limited communication capacity unveils strategies for human interaction". Sci Rep., Vol. 3, pp. 19-50, 2013.

81 de Montjoye Y, Smoreda Z, Trinquart R, Ziemlicki C, Blondel V, "D4D-Senegal: The Second Mobile Phone Data for Development Challenge", ArXiv:1407.4885, 2014

82 Telecom Italia big data challenge, [Online]. Available: http://www.telecomitalia.com /tit/en/bigdatachallenge/contest.html, 2014.

83 Bertolucci J, IBM's Predictions: 6 Big Data Trends in 2014. [Online]. Available at: http://www.informationweek.com/big-data/big-data-analytics/ibms-predictions-6-big-data-trends-in-2014-/d/d-id/1113118?piddl_msgid=196102#msg_196102, 2014.